DOCUMENT RESUME

ED 401 289 TM 025 791

AUTHOR Falk, Beverly; And Others
TITLE Issues in Designing a Learner-Centered Assessment System in New York State: Balancing Reliability with Flexibility, Authenticity, and Consequential Validity.
PUB DATE 11 Apr 96
NOTE 31p.; Paper presented at the Annual Meeting of the American Educational Research Association (New York, NY, April 8-12, 1996).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; Accountability; *Educational Assessment; Educational Improvement; Elementary Secondary Education; Instructional Leadership; Performance Based Assessment; Pilot Projects; Program Development; Scoring; Standards; *State Programs; Student Evaluation; *Systems Development; *Test Construction; *Testing Programs; Test Reliability; Test Use
IDENTIFIERS *Authentic Assessment; Authenticity; Learner Centered Instruction; *New York

ABSTRACT

The National Center for Restructuring Education, Schools, and Teaching is working with the New York State Education Department to develop an assessment system that addresses both instructional and accountability concerns. The new system is intended to move the state to a system of performance assessments used in the service of ongoing teaching and learning. The principles that govern the assessment system redesign are: (1) curriculum, instruction, and assessment must be interrelated to support student learning; (2) assessments must measure student achievement of defined standards for learning; (3) multiple forms of evidence of student learning must be consulted; (4) the assessment system should articulate standards without demanding standardization; (5) the system should be built on local involvement; (6) innovators of the system should lead; (7) support needs to be provided to teachers and schools; and (8) schools performance should not be judged solely on the basis of student outcomes. Under the redesign, each examination will include an on-demand test and a common curriculum-embedded extended task. Both components will be evaluated by a common scoring procedure. Moving the assessment system from the idea stage into operational form requires the consideration of many factors, but the first state assessments have been developed in pilot form and were tested in 1995. Preliminary findings are encouraging. An attachment lists the assessment's guiding principles. (Contains 35 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
BEVERLY FALK
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

ED 401 289

WORKING PAPER

for the symposium:

Issues in Designing a Learner-Centered Assessment System in New York State: Balancing Reliability with Flexibility, Authenticity, and Consequential Validity

presented at the
Annual Meeting of the American Educational Research Association
New York City, April 11, 1996

by Beverly Falk, Raymond Pecheone, and Linda Darling-Hammond
National Center for Restructuring Education, Schools, and Teaching

Introduction

This working paper from the National Center for Restructuring Education, Schools, and Teaching represents the collective development of ideas over the course of several years. Many people have contributed to the evolution of our thinking. We are especially grateful to Tom Sobol, former New York State Commissioner of Education and Linda Darling-Hammond, Chairperson of the New York State Curriculum and Assessment Council for originating these conceptions. We also wish to thank our colleagues at the New York State Education Department and the Cayuga-Onondaga BOCES who have been working tirelessly to enact these ideas and to make them a reality for all the students of New York State.

This paper is intended to discuss issues pertaining to work that is ongoing and continually evolving. The ideas expressed represent the views of the authors. Policies associated with these issues are the responsibility of the New York State Board of Regents and the Commissioner of Education.

Background

Increasingly, over the last several decades, educators, policy makers, and the public have come to understand that the schools in our nation must improve the ways in which we prepare students for the demands of the 21st century. A major goal is to support schooling that will encourage *all* students to construct, integrate, and apply their knowledge; to think critically and invent solutions to problems; and to respond creatively to unforeseeable issues that will confront them in the complex world of tomorrow.

In pursuit of this goal of educational excellence for a greater range of students than ever before, local, state, and national efforts have focused on articulating rigorous standards of student achievement; developing challenging curricula based on these standards that are responsive to the differing perspectives of diverse populations; building the capacities of teachers to use a range of strategies that will help students to achieve the standards; and designing and using new forms of assessment that better support and reflect what is being taught (Darling-Hammond and Wise, 1985; Mitchell, 1992; Rothman, 1995)

TM025791

BEST COPY AVAILABLE

Changes in curriculum and assessment practices have been successfully undertaken in many classrooms. Many teachers and schools have made great progress in developing learning experiences that provide opportunities for students to use and develop higher-order skills and complex problem-solving abilities. To allow students to demonstrate these learnings, they have also developed a new genre of assessments. These ask students to express ideas in deeper and fuller ways than traditional approaches have allowed. They involve students in inquiry and library research projects, compiling portfolios, keeping journals, producing videos and oral tape recordings, and conducting experiments (Darling-Hammond, et al., 1993).

This kind of teaching and assessment has remained limited in its development however, because its approach is generally not compatible with the approach to teaching needed to prepare students for the tests used by most districts and states for accountability purposes (Falk, 1995; Moss, 1995; Wiggens, 1993). These large-scale testing systems, used to make decisions such as graduation, grade retention, and placement in special programs or tracks, frequently inhibit many teachers and schools from providing opportunities for meaningful learning (Shepard and Smith, 1988; Koretz, 1988; Smith et al., 1986; Darling-Hammond, 1991, 1992; Allington and McGill-Franzen, 1992). Because the tests rely heavily on multiple-choice questions that ask students to recall facts from large bodies of knowledge, they constrict curriculum and instruction to a focus on superficial content coverage, discouraging students and teachers from pursuing more challenging and rigorous in-depth study. They drive instruction in ways that mimic not only the content but also the format and cognitive demands of tests, imposing low cognitive demands on students at the expense of applying their knowledge in purposeful contexts and on developing their abilities to creatively problem-solve (Darling-Hammond, Ancess, and Falk, 1995; Glaser & Silver, 1994; Wiggens, 1989, 1993). In addition, pressure to teach to the tests also makes it difficult to effectively teach diverse groupings of students who bring different starting points, understandings, and styles of learning to the learning enterprise (Darling-Hammond, 1989, 1991, 1994; Garcia and Pearson, 1993; Oakes, 1985).

In order to redefine this pressure between what is increasingly being acknowledged as effective classroom-based practices and assessments and the limitations of assessments in large-scale systems to support and demonstrate these kinds of learnings, the National Center for Restructuring Education, Schools, and Teaching (NCREST) is working with the New York State Education Department to develop an assessment system that addresses both instructional and accountability concerns. The new design of the assessment system is intended to move the State

from a testing program that focuses on summative evaluation using primarily multiple choice forms of testing to a system of performance assessments used in the service of ongoing teaching and learning. The new system emphasizes the integration of assessment with curriculum and instruction. This integration is intended to encourage the use of teaching and assessment strategies that help teachers, parents, and students gain a rich understanding of what students know and can do, as well as how they think and learn. The system is using a variety of assessment strategies to enable students to capitalize on their strengths, while also challenging them to perform and communicate in different ways. It will allow educators to evaluate the success of programs, while providing accountability data that indicate attainments across the State and within districts. It will help schools evaluate student learning while increasing students' opportunities to learn challenging material at high levels.

In this paper we outline the principles guiding the construction of this new assessment system. We describe the design of the system as well as the plan for how to build the infrastructure to ensure that the system gets enacted. We present findings from the first year's pilot assessments and discuss the challenges that still need to be addressed in order to construct a system that meets acceptable standards of reliability and validity for accountability purposes while maintaining support for meaningful learning for all students across New York State.

Principles Governing the New York State Assessment System Design

The assessment system redesign in New York State is guided by a set of principles aimed at supporting meaningful student learning:

1. Curriculum, instruction and assessment must be interrelated and interconnected so as to support meaningful student learning.

Anyone familiar with schools and school life knows the powerful hold that tests have on curriculum, instruction, and learning. What is tested is what is valued. Unfortunately, the converse is not always true. What is valued is not always what is tested, especially in recent decades when advances in cognitive research have revealed new information about how people learn. We have come to understand that learners actively construct knowledge through real-life experiences. Contrary to many approaches to curriculum that assume "basic" skills precede thinking skills, learners come to understand information and develop concepts while *in the course* of accruing basic facts and skills. Conceptual learning and higher-order thinking are the foundations that make all other kinds of learning possible. The implications of

this understanding for teaching is that teachers must help students to develop concepts and deep understandings of ideas and relationships as a basis for factual learning (Piaget & Inhelder, 1970; Resnick, 1987; Sternberg, 1985).

In addition, we have also come to understand that individual students learn in different ways, at different rates, and from the vantage point of their different experiences (Darling-Hammond, Ancess, and Falk, 1995; Falk, MacMurdy, & Darling-Hammond, 1995; Garcia, & Pearson, 1994; Gardner, 1983; Kornhaber and Gardner, 1993). Because of this, no highly specific, predetermined curriculum can ever be equally effective for all students. To be successful at helping all students achieve, teachers must meet students where they are and create a bridge between their individual talents, interests, and experiences, and common, challenging learning goals. This means that teachers must have the knowledge, skills, resources, and flexibility to use a variety of pathways to be responsive to the students that they serve.

These understandings about learning lead us to conclude that teaching and curriculum must encourage the development of students' conceptual abilities and higher-order thinking skills - rather than feeding them bits of disconnected information - and must permit teachers to link new concepts to students' very different experiences and prior understandings - rather than assuming that students will all learn the same things in precisely the same ways. Such an approach to teaching and learning will help us to develop the kinds of learners our society needs - creative, critical thinkers who can put knowledge to use solving problems, offering alternatives, thinking deeply and in multi-dimensional ways, and working cooperatively with others.

In order to realize this vision of teaching and learning, assessment too must be transformed. If learning experiences are to be challenging, coherent, and aimed at developing the full range of students' capabilities, assessments must be designed to reveal the complexity and full range of students' learning. Assessments must be rich and dynamic, filled with information about students' potentials as well as progress, and motivating to students, teachers, and schools as it illuminates compelling goals. Support for meaningful learning therefore, necessitates the development of "authentic" assessments (Shepard, 1995; Wolf, 1989). Such assessments

- look directly at student work and performance
- measure the use of knowledge and skills in real-world contexts and applications;

•require higher-level thinking and complex problem-solving;

•are directly connected to the purposes and goals of learning;

•evaluate student progress and achievement through the use of public, open, and clearly articulated criteria that guide teaching and learning.

•provide information that is useful for instruction and that encourages reflective practice;

•examine the process as well as the product of learning so as to enable teachers to assess student growth in a cumulative, longitudinal fashion;

•provide multiple ways for students to demonstrate their knowledge, skills, and understandings about many dimensions and kinds of learning

•are easily expressed, used, and understood by students, parents, teachers, and the public;

•communicate expectations and support student motivation, self-assessment, and continual growth;

•are embedded in the ongoing life of classrooms, requiring teachers to assume new instructional as well as new assessment roles, calling on teachers - those most closely involved with and knowledgeable about students - to be evaluators rather than outside experts who have little sense of the context of the learning environment.

•are accurate and valid means for identifying students' strengths, abilities, and progress, opening up possibilities for encouraging further growth rather than precluding access to future advanced instruction.

Assessments that embody these characteristics look different in different contexts and different age levels. In early grades they focus more on the process of learning, relying heavily on observation and documentation of learners' growth over time in natural contexts. They yield information about the uniqueness of the learner and how each student is progressing in relation to developmental criteria along an articulated continuum for a discipline or age group (NAEYC, 1988). In upper grades the best assessments include some means for observing and documenting student

growth while focusing more on the products of student learning, using a range of projects, performances, and exhibitions to evaluate and assess student achievement according to articulated criteria that are important for actual performance in that field Darling-Hammond, Ancess, & Falk, 1995; McDonald, et al., 1993; Mitchell, 1992; Perrone, 1991).

Both of these types of assessment rely heavily on teachers' use of assessment strategies that are embedded in the curriculum. Much like the assessments that are used predominantly in other countries, they include extended tasks and projects which call on students to analyze, investigate, experiment, cooperate, and present their findings in written, oral or graphic ways as well as portfolio collections of student work in selected subject areas. The information they provide - understandings of students' strengths and approaches to learning - is useful to teaching and helps teachers shape and adapt instruction to the needs of individual students (Darling-Hammond, Ancess, and Falk, 1995; Wood and Einbender, 1995).

2. <u>An assessment system must be designed to measure student achievement of standards for learning that delineate what students should know and be able to do as a result of their education</u>.

Assessments designed to measure student of achievement of standards are very different than the norm-referenced multiple choice tests still in use in most large-scale testing programs. Norm-referenced tests are designed to rank order students on the basis of achievement in a particular subject area. Items on norm-referenced tests are designed and purposefully selected to ensure that student scores will fall out in a bell curve, i.e., with a majority of students attaining scores within a mid-range and only a small percentage of students attaining scores in the upper or lower percentiles. While they try to level the playing field through uniform administration requirements - such as multiple choice responses, time restrictions, and limits on the use of resources during test-taking - these "one-size-fits-all" conditions and contexts result in constraints on the authenticity of the items and limits to students' ability to demonstrate what they know and can do. This deliberate construction of the test to sort and track students is not compatible with the goal of supporting *all* students to achieve high standards (Darling-Hammond, 1994; Oakes, 1985).

In contrast, all items or tasks on standards-based assessments are developed to provide indicators of student attainment of valued learning goals. While student performance on them can vary greatly, they are not designed to preclude some

students from achieving mastery. Variation of student performance results from genuine differences in what students know and are able to do.

In addition to breaking out of the norm-referenced mode, one of the advantages of standards-based assessments is that they contribute to supporting clarity of purpose and rigor of curriculum (Darling-Hammond and Wise, 1985; Resnick, 1989). They ask all of those involved in educating students to be clear and purposeful about articulating their goals. They call on teachers and schools to develop curriculum and instruction that focuses on the essentials of learning defined by the standards. They push all those involved in curriculum decisions to ask and answer the fundamental question: "What will students really know and actually be able to do as a consequence of engaging in particular learning activities?"

Standards-based assessments offer all students the opportunity to demonstrate how they have mastered valued goals for learning in a particular discipline or domain for a particular age group or developmental level. Because they start with learning goals and end with performance standards, providing descriptions of varying levels of achievement, they allow students to demonstrate a range of abilities, from beginning competencies to distinguished performance.

3. An assessment system should provide multiple forms of evidence of student learning for multiple purposes - but all components of the system should always support student learning. While each individual authentic assessment should strive to provide opportunities for students to demonstrate in a variety of ways what they know and are able to do, a system of assessments should have multiple forms of evidence about student progress and achievement that can be used for multiple purposes (Price, Schwabacher, and Chittenden, 1993). All of these forms of evidence should be consistent with teaching that fosters meaningful learning. Test preparation, in other words, should always mirror good instruction (National Forum on Assessment., 1995).

Assessments in the system should complement each other and be able to serve multiple purposes. The primary purpose of any assessment system should be to provide information that can be useful to teaching. Assessment systems should yield evidence about what students know and can do, the strategies students use, as well as what strengths, interests, and needs they bring with them to the learning situation. Knowledge of this can help teachers to teach - to diagnose student needs and adapt their teaching strategies to be responsive to students' individual differences.

A second purpose of an assessment system should be to demonstrate student achievement. Assessments should provide teachers, students, families, and school

systems with information about how students have progressed over time, what standards for learning they have already mastered, and what further challenges they need to take on to progress through the continuum of achievement.

A third purpose of assessment systems is to furnish information about the progress of groups of students across schools, districts, or states so as to keep track of how students from different backgrounds are doing across different locales and across populations. Such information will help educational agencies ensure that they are providing comparable and equitable learning opportunities and thus are fulfilling the public trust placed in them to educate all students equitably.

As it is now, in order to fulfill these different purposes for assessment, teachers often have to prepare for and subject students to extensive testing that give conflicting messages about the kind of learning that is valued. For example, in some situations teachers are collecting student work and preparing their students to present projects and exhibitions that offer a richer picture of student progress than any one test can ever provide. Yet this evidence of student learning has no way to be used to document achievement for accountability purposes.

In designing systems of assessment, classroom-based assessments that are useful for teaching should be able to be used for accountability purposes too. The assessment system should use multiple forms of evidence to make important decisions about a student's future. It should clearly define who is responsible for what decisions, with those who are most knowledgeable and closest to students and their learning - teachers - being held responsible for making the decisions that affect students' futures.

4. <u>An assessment system should articulate standards without demanding standardization.</u>

An assessment system should provide a clear direction for teaching and learning (accomplished through the articulation of standards) but allow flexibility for local interpretation in the teaching and assessment of the standards. The standards of the system should be uniform but how the standards are taught and demonstrated can vary. No one teaching method or strategy is equally effective for all students. Differences in students' styles, approaches, and paces of learning require that students be offered a variety of different entry points and paths to learning. Flexibility needs to be provided in the context as well as the content of the studied discipline. Does it matter if the principles of physics are learned through designing a racing car or through calculating the speeds of different amusement park rides? Cannot the same

understandings about the elements of narrative style be obtained from reading and writing a variety of authors?

In the same way, students' varying backgrounds and experiences should be taken into account when assessing their understandings and mastery of valued skills. Students from different cultures and locales, who speak different languages and have different customs, will both learn and be able to demonstrate their learning in a variety of different contexts. For example, students who live in coastal areas might learn about the properties of water in a different manner than those living in the midst of a mountain range. Students in urban areas might engage in very different social studies projects than those who live in rural areas. All of these students can learn the same concepts and gain the same skills if allowed to acquire them through routes that connect with their backgrounds and experience.

One of the major challenges in developing large scale assessment systems is to find a way to be flexible and responsive to the differences that exist across and among large groups so as to be fair and to support genuine learning, while still providing comparable and reliable evidence for public accountability purposes Darling-Hammond, Ancess, and Falk, 1994; Falk, 1995; Moss, 1995; Wiggens, 1993). Serving these dual purposes simultaneously creates a tension that is not easily solvable. Part of the tension has to do with the fact that in order to use assessments to compare student achievement across large groups, a certain amount of standardization of the assessment must take place. This standardization is necessary to assure that the student work being compared is measuring the same standards of learning and represents the same quality of achievement across sites (an issue that does not arise when using standardized tests in multiple-choice formats where achievement is easily measurable because there is only one possible right answer).

Efforts to make authentic assessments become more standardized to answer reliability and comparability concerns generally affect their validity and usefulness for teaching. Standardization, by necessity, makes assessments become less contextualized and less connected to individual students (Darling-Hammond, Ancess, and Falk, 1995; Moss, 1995; Wiggens, 1993). So in developing a system of standards-based assessments a way must be found to design assessments that are responsive to variations in students and their different learning contexts, while also putting a mechanism in place to assure that the assessments and their scores elicit and represent the same quality and quantity of achievement. This can be accomplished by designing assessments for the system based directly on uniform standards and that are scored through standardized scoring processes. Content,

context, procedures, and format can all vary according to the precise learning situation. Standards and rubrics are uniform, creating a validity and reliability path from standards to tasks to student scores.

5. The assessment system should be built on local involvement.

Who is involved in assessment design is as important as what assessment tools and strategies are used. The development of an assessment system that is aimed at supporting teaching and learning must be informed by teachers and grounded in work at the school level. Assessments then should be developed, implemented, and scored by teachers. Involving teachers in evaluating work and discussing standards provides them with information that can be used to support instruction. It also builds teachers' professional knowledge. From first-hand encounters with assessment development and evaluation of student work, teachers learn about the deeper structures of curriculum, the natures and nuances of student thinking, and the connections between teaching efforts and student performances.

Assessments that are externally developed and scored cannot transform the knowledge and understandings of teachers and of school organizations in this way, even if they are more performance-based than are current tests. At the core of an assessment system then, is the principle that assessment should inform and support teachers' efforts to understand student learning and schools' efforts to improve the educational opportunities they provide.

Local involvement entails more however, than just teacher input. Local involvement must also include students, their families, and community members. These voices need also to be reflected in assessment systems - through participation in standard development and/or evaluating student work, as well as through involvement in two-way communication processes that provide information as well as feedback about student progress and school performance. Assessment systems must develop mechanisms for regular dialogue and exchange of ideas between all concerned with the educational enterprise. Methods for communicating to parents and the public should include exhibitions of students' work, such as displays and demonstrations of what they have learned and accomplished. As in the town meetings and recitations conducted at schools in the early days of this country, people should be able to see what schools are doing and what their students can do as a result. These processes of communication should be collaborative, providing opportunities for the community to participate.

6. Let the innovators of the system lead.

Local innovation generally precedes most large-scale system-wide change. Changes in system-wide policies and structures usually require actions and strategies that are more lengthy and complex than local action that frequently is able to quickly and directly result in change. For this reason, those responsible for system policies should learn from local innovations and find ways to share knowledge about successful practice. The assessment system should document and disseminate promising local assessment practices so that people in the schools as well as the public can learn more about how these support student learning. To support and promote innovation, the system should also sponsor the development, synthesis and dissemination of research and information about new forms of teaching, curriculum and assessment so that widespread access to these resources is available.

Because teachers and schools, just like students, are at different stages of development and implementation of new teaching and assessment practices, the system should encourage existing innovative work. Freedom from existing policies that constrain promising experimentation should be granted to those individuals, schools or districts that have made progress in developing their own assessment systems and mechanisms for accountability. In this way, reform efforts will be complemented and not held back or unraveled by the existing system-wide policies and processes.

7. Supports need to be provided to teachers and schools to build their capacities to enact new teaching and assessment practices.

Professional learning should be an ongoing part of the process of developing and evaluating curriculum and assessments. As teachers consult with one another in collectively developing, analyzing, and evaluating student work, they learn about student learning and gain insights and understandings about their own learning processes as well. Personal experience and engagement with their own learning seems to, in turn, heighten even more their sensitivity to the complexities and issues surrounding meaningful teaching and learning.

An important way to support this learning is to provide time, along with the resources of research, information, and expertise, for teachers to work together on the development and implementation of new standards, curriculum and assessments. Providing time in the school schedule for collaboration aimed at improving teaching and learning will act as a powerful incentive for faculties to engage in the kinds of

changes necessary to provide all students with the opportunities to learn more challenging material in more authentic ways.

8. How schools are doing should not be judged solely on the basis of student outcomes. While assessment systems address issues of accountability, these cannot be fulfilled only by looking at student outcomes. Assessment of student outcomes must be accompanied by other assessments that examine school inputs and practices likely to produce the valued outcomes. Assessment systems must also include assessments that provide information about 1)how well school personnel are using professional knowledge and practices to meet the needs of their students; 2)if students have equitable access to resources and materials that support their ongoing learning; and 3)if schools have structures and vehicles in place to collectively address the problems and issues which arise in the course of teaching. Because all of these aspects of schooling critically affect how students are achieving, they must continually be monitored right alongside the assessment of student achievement (Darling-Hammond, 1992). School review processes, much like the Inspection system developed in the England, provide this kind of information about the quality of teaching and the opportunities for learning that are made available in a school. School and district reporting systems must also include this information.

Understanding the New York State Tradition

In order to fully understand the New York State assessment system's redesign, a sense of history and context is helpful. New York State has a long tradition of administering an extensive array of examinations to students. In the elementary grades these have been used for program assessment as well as for identifying students in need of special services and for making decisions about grade promotion. Aggregated scores at a school level have been used to identify schools in need of special supports.

In the high schools, state examinations since the 1940's and 50's have traditionally been highly-structured and content-laden, strongly influencing the course of programs and classes in the disciplines. The examinations have been tightly aligned to the scope and sequence of course syllabi, determining to a large extent the way in which teaching and learning takes place in the classroom. Selected teachers have traditionally been called upon to create questions for the exam and all teachers administering the exam have been primarily responsible for scoring. The State's role has been to audit the process in order to flag irregularities in the scoring.

State high school examinations in a range of disciplines have always been a prerequisite for graduation in New York State. Two types of exams and diplomas have been offered - Regents Competency Tests, leading to a locally-approved diploma, which ask students to demonstrate the minimum competency expected by the State; and Regents examinations, leading to a Regents-approved diploma (in the past considered the college-bound diploma), which require students to demonstrate mastery over more difficult content from their coursework. Both types of examinations have focused heavily on content coverage through the use of multiple-choice questions. The two types of exams, originally created to address differences in student abilities, have inadvertently created a two-tiered structure which determines access to curriculum in New York high schools. Students in Regents Competency courses do not get exposed to the same content or curriculum as students who are in the Regents track. While the Regents track is composed of more difficult content, both examination systems provide few opportunities for students to demonstrate higher order thinking skills and application of knowledge in a range of different ways.

The system's redesign hopes to remedy the discrepencies in opportunities to learn that are prevalent among students with differing abilities and backgrounds in New York State, while at the same time providing a means for all students to access and demonstrate learning in broader ways than traditional measures have allowed. Under the system's redesign, all students will be required to take the same set of examinations designed to measure attainment of the State's standards in core areas. After fulfilling these requirements students will then have the choice to pursue more advanced level courses and exams or to branch out into other areas of study. The natural range of student abilities and performance will be accounted for by allowing students to sit for the required examinations at different points in their school careers - some students can move rapidly through the required curriculum and exams to take on more demanding and challenging work while those who struggle to meet the state standards can proceed at a different pace. But all students will initially have access to the same content. The intention is to provide every student with an equal opportunity to achieve excellence.

A Framework to Support Meaningful Learning

The new system of assessment in New York will evaluate students' achievement of standards within and across seven defined curriculum areas, developed by diversified committees of educational stakeholders from across the State to guide the design of instruction and assessment. State and local assessments

will work together to serve a number of different purposes - providing information to students, teachers, parents, and others about individual students' progress in ways that help inform teaching and learning; providing information to the State and local school boards about the outcomes of programs and the performances of groups of students in ways that can inform the further development of state and local policies and practices as well as identify schools in particular need of improvement; communicating expectations and standards about valued knowledge, skills, and abilities that can support local dialogue and program development. The system design is intended to address these purposes through a small number of high-quality, state-administered assessments that *stimulate, complement, and build on local work.* This is a key point: the system strives to incorporate and value instructionally-useful, curriculum-embedded local work for accountability purposes. It strives to do this while meeting the rigorous technical psychometric requirements generally applied to large-scale, high-stakes testing systems.

Components of the State Assessment System

Under the redesign of the New York State assessment system, each examination includes both an on-demand test and a common curriculum-embedded extended task.

The on-demand test components of the examinations are performance-oriented standardized examinations administered at a given point in time. These assessments are time-bound, not timed, but can vary in duration - lasting anywhere from several hours to several days. The test component calls on students to demonstrate what they know and can do by answering questions in written or oral responses, conducting experiments or other short-term investigations, or producing products ranging from written essays to graphs, charts, and computer simulations.

Curriculum-embedded extended task components of the examinations include classroom projects that are part of the ongoing teaching/learning process: collections of written products, artwork, designs, or other exhibitions; extended investigations or experiments; research projects; or individual and collaborative group work. These kinds of assessments can be conducted at various points in the year as teachers deem appropriate, taking anywhere from several days to a semester to complete. In addition to their use for state assessment purposes, the projects in the extended task component of the state assessments can form the core set of tasks in a K-12 collection of student work, that local districts can use to assess individual student progress and to inform the teaching process.

Each of these components are constructed to directly assess student achievement of State standards that have been articulated for each discipline. The scores students receive on each component are combined to determine each student's score on the examination as a whole. The information about learning that these exams provide are intended to be useful for informing the public about how students and schools are progressing toward attainment of the standards.

In addition to these state examinations, local districts will develop their own set of complimentary assessments (performance tasks and portfolio collections of student work) that will document students' continuous progress and will inform teaching and support learning. These will include collections of student work that illustrate student abilities, thinking, accomplishments, and approaches to learning; other learning records based on teacher observations and documentations of student learning in natural classroom contexts; curriculum-embedded assessments of student progress such as performance tasks and extended projects that are part of the teaching/learning process; and on-demand performance tasks and tests that pose common requirements of students. This work will lend itself to the development of local K-12 portfolios, in which the extended task component of the State exams can also be kept. Districts and schools that have made progress in developing and using performance-oriented assessments can fold the curriculum-embedded component of State assessments into their existing local systems. In districts and schools where performance-oriented assessment systems have not as yet been developed, local assessments can be built on the models provided by the state examinations' curriculum-embedded component.

To assist schools and districts in developing such assessments, the State is compiling an Assessment Collection that includes prototypes of promising practices developed in the field. Technical assistance and professional development will be needed to assist districts in developing performance-based assessment systems for local accountability and curriculum reform and to assist districts to align these local assessment procedures to the State standards.

Scoring: Establishing Reliability and Validity

Both the "test" and the "extended task" components of the State assessments are evaluated by a common scoring process mapped directly to the standards outlined for each discipline. Examination scores will be a composite of the scores earned on the test and the extended task components. Scores will be differentiated into multiple levels of performance (for example, beginning, proficient, accomplished and

distinguished). A proficiency standard will illustrate the minimum score that a student must achieve to be considered competent enough to pass; an accomplished standard will be set to identify performances that represent a strong command of the subject and skills being evaluated; and a distinguished standard of accomplishment will be identified to acknowledge exceptional performance.

The scoring process will be standardized through the use of common rubrics based on the State standards for each discipline. Evaluating work with these rubrics will allow local teachers and schools to flexibly develop and use curriculum-embedded projects and collections of student work while at the same time ensuring that the work can be assessed in a comparable manner. All teachers who score performance tasks will participate in scoring protocols and moderation exercises so that common understandings can be developed about what constitutes achievement of the standards. These processes of moderation and auditing will be used to establish the reliability of teacher ratings of student performance. Common rubrics and the scoring processes around them also promote shared understandings about curriculum and teaching, contributing to the knowledge base that constitutes teaching as a profession.

The validity of the State examinations is established by evaluating whether they are good measures of the knowledge and skills embodied in the State standards. Validity also rests on the premise that standards are understood by teachers and that they shape their curriculum (that is, provide the necessary opportunities to learn) to meet the standards. Demonstration of consequential validity will ultimately require a showing that the use of the assessments encourages the kinds of learning that the frameworks intend.

Raising Standards and Providing Flexibility

Including performance elements as part of the test component of the Regents examinations as well as requiring a curriculum-embedded extended task component will raise standards and ask students to take on more challenging work than the current system. Adding performance activities to the examinations taps into higher level skills - such as synthesis, analysis, evaluation, problem solving, extended writing and research - that cannot be measured reliably or validly through most current assessment practices. Research on testing indicates that the addition of performance components on examinations increases the difficulty of the examination and raises standards.

Standards will also be raised at the high school level by eliminating the dual examination system and replacing it with a unitary set of exams that all students will be required to take. Merging the two systems will require all students to undertake more intellectually rigorous tasks and will not track some students out of challenging courses solely because of the structure of the testing system. The expectations of the new Regents examinations -- for students to demonstrate not only attainment of knowledge but, additionally, the use and application of knowledge -- will apply to all students. However, performance levels on these more challenging tasks will vary, enabling more advanced students to demonstrate very high levels of attainment, while all are challenged to undertake more authentic and challenging performances.

Flexibility in the programs will be achieved through expanding the format of the Regents examinations to incorporate extended tasks as a key component and by tying examinations to standards rather than to specific course configurations. (Thus, local schools or districts may choose to use a variety of course configurations that prepare students to succeed on the assessments.)

The extended task component will allow students to work on in-depth curricular tasks that are designed to elicit critical thinking and to demonstrate a student's command of major concepts in a particular field of study. The extended task component of the Regents examination may take the form of an "exhibition" - a student demonstration of long term work in an area of study which utilizes a variety of presentation modes. For example, student exhibitions may include a written component, an oral presentation, an authentic representation of performance (e.g., art portfolio, dance, technical design).

The extended task components will vary in their structure: they may take several days, weeks, or longer to complete. Some may be structured as a culminating activity of work completed over the course of multiple years within or across disciplines. For example, the extended task portion of the English examination might include a portfolio of written work assembled across more than one year and more than one area of study, including an historical research paper, for example, as well as literary and artistic critiques. Similarly, a project that integrates math, science and technology might be submitted for Regents credit to supplement on-demand examinations, but the project might <u>not</u> be connected solely to one particular course. Rather students might work on an experiment or exhibition across several discipline-based or interdisciplinary courses in mathematics, science, and technology and/or over more than one year.

Such a project would be in addition to completing the test component of a Regents examination in mathematics that evaluates the attainment of content and performance standards represented in the new framework. This test, like the College Board's mathematics achievement tests, would be taken by students who could have had different configurations of mathematics courses, so long as they included the core concepts and skills being assessed.

Using this project-based approach provides for flexibility in how districts develop curriculum and course structures. It allows districts to have the flexibility to make decisions about the structure of course sequences and the content of coursework, for example, whether they will offer discipline-based or interdisciplinary courses. For example, preparing students to meet the standards for the integration of math, science and technology may be accomplished through a traditional disciplinary approach to science that maintains the separation of the different sciences (e.g., general science, biology, chemistry) or through an integrated approach to science that combines the different science disciplines.

To accommodate these changes, all new examinations will be phased in over a three to five year time frame. This will provide districts with time to align their curriculum to the state-adopted standards and to develop structures that will provide all students with the opportunity to meet the new standards.

Developing and Piloting the New Assessment System

Moving this assessment system from the idea stage into operational form requires consideration of many complex factors. How do you "unfreeze" a system to be open to change and retool it for new purposes and practices? How do you inform, educate, and win the confidence and support of the public about the changes that are taking place? How do you build the capacity of both the State Education Department as well as of teachers and other educators in the field to understand and use a new assessment system? How do you create new infrastructures and provide the resources and supports that are needed to ensure that the change initiative takes root and stays alive in the face of budgetary constraints and political pressures?

Strategies for Unfreezing the System

In order to create an atmosphere that invites change, a variety of strategies have been initiated to "unfreeze" the old system. One strategy with which we began was to permit those "leading-edge" schools and districts, who have been using student-centered pedagogy and practices for some time, to be freed from the

constraints of existing policies so that they could experiment and innovate with new forms of assessment. Waivers from existing tests were granted so that these schools and districts could have the flexibility to develop alternatives that would both complement and evaluate their innovative work. Other strategies were also initiated with the intention of encouraging experimentation: the Regents Options Project (an initiative that provided professional development supports for interested teachers, schools, and districts to develop curriculum-embedded assessments that could be credited up to 35% of the existing Regents examination score); State participation in the New Standards Project pilot examinations and portfolios; the establishment of New York State "Partnerships Schools" - comprehensive waivers awarded to schools and districts that have created fully-developed alternative assessment systems. As a result of these "invitations to invention" offered by the State, a climate and culture of systematic experimentation has developed. Through involvement in these experimental initiatives, both individuals and organizations have increased their capacities to teach and assess in new ways.

Constructing Prototype Exams

Much of the work resulting from this proliferation of experimentation is now being used to inform and guide the construction of the performance-based exams that comprise the new State assessment system. The first phase of this process has been to construct full-blown prototype examinations which are to be initially piloted on a small scale. The rationale for developing full-blown prototypes as opposed to a more incremental approach to changing the examinations is that prototypes allow the State Education Department to signal the field and the public about the direction that the redesign effort is taking. Prototypes make it possible for a full image of the new exams to be presented - from their new philosophical underpinnings to their new form and content. Piloting these exams initially on a small scale provides an opportunity to gather preliminary information about the feasibility of the exams as a whole - the content validity of the tasks, the reliability of scoring rubrics, the length and difficulty level of the exams in their entirety, the impact of the exams on students, teachers, and schools. This information will then get used to revise and refine the examinations for a larger second year pilot which will subsequently be used to begin another cycle of revision and refinement. Through such an interative process over several years it is anticipated that the exams will be honed to become as effective and useful as possible. Third year pilots will be on a sufficiently large scale to gather data for official reliability and validity studies. After ensuring that the exams have the technical

capacity to serve as high-stakes accountability measures and that they also have the substance to be meaningful measures of genuine learning, the exams are planned to go "on-line" for students across the State in the fourth year of the development process.

Alternative Equivalent Measures

To ensure that the new system, once it is established, does not supplant ongoing innovation, a mechanism will be put in place to encourage continual innovation, renewal, and growth. The new assessment system will continue to provide a means for some assessments or assessment systems to receive waivers from the established system if the alternatives can be demonstrated to be of at least equivalent rigor to the existing state examinations. The College Board's Pacesetter or Advanced Placement examinations, the International Baccalaureate exams, or other locally-developed and state-approved assessments might be used as substitutes for State-developed examinations.

Decisions to use alternative assessments for state credit should be based on a demonstration that the alternative meets or exceeds standards established for the exam it is designed to replace and that the alternative assessment is aligned to the standards of the curriculum framework in the designated subject area. Decisions about the appropriateness of alternative examinations need to be made by a State-convened established committee of recognized experts in the fields of curriculum and assessment.

Capacity-Building to Prepare for the New Assessments

Building on a long-established New York State tradition of involving teachers in creating and scoring assessments, construction of the new system of performance-based assessments is being done by a collaboration of teachers (who represent different geographic locations and the diverse populations of the State), State Education Department staff, and consultants with expertise in performance assessment. Their work is informed by promising work that has been developed by other State and national assessment development projects. All of these combined inputs help to ensure that the new assessments reflect the most current theory and practice in the profession.

Teacher involvement is also being built into the scoring procedures for the new assessments. Unlike multiple choice scoring, which calls on teachers to discriminate only between right and wrong answers, scoring of open-ended performance-based

assessments calls on teachers to assess problems that, in most instances, have multiple right answers. Thus the State needs to provide teachers with opportunities to learn how to score these new assessments. In the course of learning how to score, teachers will be asked to look carefully at student work and, as a consequence, will gain information about students and their learning - information that they can use to shape their teaching in ways that are responsive to their students and their students' needs.

Assessing student work against the new standards also helps teachers to become clear about the important dimensions of their discipline and the criteria for performance that constitutes excellent. As teachers dialogue together about student work and about how the work reflects the standards, they create shared understandings and a common language. This results in a strengthened sense of professionalism as well as a greater probability that the scores they assign to student work will concur with each other, thus enhancing inter-rater reliability ratings.

Another way that capacity is being built for the new assessment system is by circulating the prototype pilot assessments, along with scoring rubrics and student work samples representing different levels of achievement of the standards, to every district in New York State. Study and discussion of the new assessment prototypes provides teachers, administrators, parents and the public with the opportunity to learn about the new assessments themselves as well as the educational philosophy they represent. Public awareness can serve to enhance pedagogical understandings as well as build public support for the desired change.

Creating an Infrastructure to Scale-Up the Work

Given the fact that the new New York State system of assessments is being constructed to include and value teacher judgment of student work embedded in the contexts of local curricula, it is essential that the design of the system include time, resources, and mechanisms for teachers to learn about performance assessment design and scoring. An infrastructure is thus being created to teach a State-wide core group of lead teachers about the nature of performance assessments and about how to construct them, create criteria and scoring rubrics for them, score them, and teach others to do the same. Initially, for the purposes of technical scoring studies, lead teachers will do the official scoring. They will then return to their regional centers to lead other teachers in scoring local student work. It is anticipated that regional centers will also become the locus of these and other professional development activities for the assessment system in the State.

First Year Findings

The first State assessments to be developed in pilot form were in the disciplines of mathematics and English language arts. In spring 1995, seven elementary and high school pilot assessments were launched among a small but representative sample of the State. (This year the assessments have been refined and are being piloted among a larger sample.) Preliminary findings from the first pilot year are encouraging. They indicate that the tasks developed for the exams offered reliable and valid ways for students to demonstrate their achievement of the challenging learning goals articulated in the State standards while also providing information that is useful for teachers to use in instruction. Interrater reliabilities of over .9 for congruent scores were achieved in mathematics and of .7 were achieved in the language arts. An overwhelming majority of the 70 teachers who administered the pilots felt that the examinations were accurate and fair measures of the standards. Many reported that the exam gave them more information about their students than traditional tests do:

> Students were able to show me what they know instead of bubbling in an answer that they guessed.
>
> The really important part of the test was that the kids showed the work. It made me see how they were thinking.

Many teachers also reported that the exams supported needed changes in their teaching:

> "The exam serves as a guideline for teachers."
>
> "The assessment requires teachers to teach differently."
>
> The exam gave me a look at my students but also at myself as a teacher. I thought I was teaching math before but this test made me realize that I need to make things more real and practical. It gave me a chance to see what I need to do in my teaching.
>
> The test really modeled good practice!

A majority of teachers also reported that their students enjoyed taking the exams. Some of their comments about this were:

> "Most of my students enjoyed the assessment because they believed that this was a better instrument to measure what they know."

"The students really enjoyed this. I think they liked being able to have as much time as they needed. Most importantly, I think they enjoyed it because they could really relate to the questions."

"My students were eager to do the assessment and enjoyed the challenge."

"This was an engaging activity. It was an authentic task that got the students involved."

The test allowed the struggling kids to do *something* while the "most intelligent" child in the class also found it challenging because in order to do well he had to think. It provided an opportunity for all students to show us their best work.

The kids had a great time. They learned a great deal. They changed over the days that they did the project. It was a wonderful experience!

Students too responded positively to the new content and format of the tests. They reported that they liked the new exams because the exams assessed their skills and abilities better, allowed them to express their ideas more fully, were challenging and engaging, related to their lives, gave them enough time to show what they could do, and provided them with a learning experience. Some of their comments about the exams were:

"I like this test, and I think it's better because it shows more what you are."

"You can explain what you think instead of what *they* think."

"I get to actually write and spell instead of filling in those blanks. Filling in the blanks is not helping you to read and write."

"I liked all parts of the test because they were all interesting."

"It was hard and fun and it gave me a sense of accomplishment."

"This was a good challenge to your ability!"

"It is interesting to learn while you are doing a test."

Next Steps and Challenges

One of the difficult part of any design is developing details to match the design conception. Many details still need to be addressed before the assessment system that we have presented here can be fully operationalized. Some of the challenges that remain to be met are:

Challenge 1: Managing the transition from the old to the new assessment system

Given the comprehensiveness of the New York Assessment system, a four year phase in strategy is being adopted. More specifically, each new assessment is designed and developed by teachers over a three year period. Year 1 is designed to experiment with new testing formats (such as curriculum-embedded projects, portfolios, student learning records) and to examine the feasibility of these approaches for large scale assessment. Year 2 is dedicated to further experimentation and refinement of the Year 1 assessments tasks. In Year 3 a large-scale representative sample is drawn to develop multiple assessment forms and to establish reliability and validity for the assessment. Year 4 is an-line year to implement the new assessments. In each of Years 1-4 scoring rubrics are developed to both build the capacity of a critical mass of teachers statewide to use the rubrics and to obtain estimates of the reliability of the scoring system for a statewide administration.

Challenge #2: Solving technical measurement issues

The use of performance assessment for a high stakes accountability system poses some unique measurement dilemmas:

a)How to aggregate data sources using multiple forms of evidence;

b)How to develop reliability of teacher scoring;

c)How to ensure comparability of the assessments from year to year;

d)How to measure student growth over time.

We also need to grapple with how to report the scores of the assessments in a meaningful and publicly-acceptable way. Other more philosophical and foundational problems that continue to be relevant and not sufficiently resolved have to do with issues of accountability: What is the appropriate unit of analysis for accountability - the school or the student? Does there really need to be universal testing or would sampling of students suffice?

Challenge #3: Moving the change initiative beyond a small scale prototype to a large scale system

Still to be faced are additional issues that relate to the process of enacting change. Scaling up requires that the entire old system be retooled to meet the demands of the new assessments - from the test developers to the teachers in classrooms. At the development end, test developers must be supported and resourced to build their capacity to produce an ongoing bank of high quality assessments. This entails exposing them to new information and research, new

models of assessments, and other colleagues who are engaged in pioneering work in the field. Retooling the system also entails expanding on the pool of teachers who are involved in creating the assessments to ensure that the teachers adequately reflect and understand the backgrounds and abilities of the State's diverse student population. Launching a Teacher-in-Residence program that allows practicing teachers to bring their expertise to the service of this initiative would be one cost-effective way to guarantee ongoing participation of such teachers. Establishing a technical advisory group for the review of all new assessments is another strategy that will help to bring the most up-to-date and expert knowledge to the development process.

At the classroom end teachers must be supported to teach in ways that the new assessments model and mirror. Teachers must also learn how to score assessments reliably and how to use the information about student learning that the assessments provide to inform and enhance their teaching. The State Education Department must find new ways to provide resources for this purpose and to support local districts to do the same.

Challenge #4: Evaluating the impact of the new assessments

Is the new assessment system producing the types of effects on curriculum, instruction, and learning that were originally intended? This is a key dimension of validity and needs to be systematically addressed during the implementation of the new assessments and beyond. We are developing a research agenda to chronicle the effects of the new system on teaching, learning, and student performance as well as to getting a reading on perceptions about the assessments from those individuals who are involved. Through the research we hope to find out: if performance assessments do indeed allow a greater range of students to demonstrate their knowledge in a wider range of ways; how performance assessments effect teaching; views of teachers, students, and parents about the impact of performance assessments; how student performance on these new assessments compares with student performance on more traditional standardized tests.

Toward A System that Supports Meaningful Learning

As understandings continue to increase about the kind of education that supports a greater range of citizens to achieve at higher levels of performance, all educators are challenged to also expand our definition of what constitutes sound assessment. We need to develop assessments that do not constrain and constrict

teaching and learning. We need to develop assessments that support what we have come to know is needed for learning: opportunities to use and apply knowledge, to inquire, to analyze, to critically evaluate, and to use creativity to pose and solve problems. We need to make these assessments an integral part of the learning experience, allowing students to demonstrate in a variety of ways, suitable to their individual talents, what they really know and can do. Only when real and meaningful student work is made a part of the assessment process can there be valid and equitable evaluation of the skills and abilities of all students.

The redesign of the New York State assessment system offers one model of how to enact these principles and policies. The beginnings of the system that are described here promise to provide opportunities for students to demonstrate what they know and can do in both standardized and non-standardized settings. The new assessment system answers the need to provide information about student learning that is valid and useful for teaching and learning as well as for reporting the outcomes of teaching (student achievement across large groupings) to the public for accountability purposes. It establishes a basis for generalizing about student performance and achievement within and across large and diverse groupings. In addition, it is motivating and enjoyable to both teachers and students, a compelling reason to continue this kind of development. Perhaps future assessment development work in New York and elsewhere in the country should heed the unsolicited advice of one of the students who participated in taking the pilot examination:

> "This test is better than other tests. You should make more tests in the future just like it."

References

Allington, Richard L. and Anne McGill-Franzen (1992). Unintended effects of educational reform in New York, *Educational Policy, 6* (4): 397-414.

Darling-Hammond, L. (1989). Curiouser and curiouser: Alice in testingland. *Rethinking Schools, 3*(2), 1, 17.

Darling-Hammond, L. (1991). The implications of testing policy for educational quality and equality, *Phi Delta Kappan, 73*, 3, 220-225.

Darling-Hammond, L. (1992). *Standards of practice for learner-centered schools.* New York: National Center for Restructuring Education, Schools, and Teaching (NCREST).

Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review, 54* (1), 5-30.

Darling-Hammond, L., Ancess, J., & Falk, B. (1995). *Authentic assessment in action.* New York: Teachers College Press.

Darling-Hammond, L., Einbender, L., Frelow, F., & Ley-King, J. (1993). *Authentic assessment in practice: A collection of portfolios, performance tasks, exhibitions, and documentation.* New York: NCREST.

Darling-Hammond, L. and Wise, A. (1985). Beyond standardization: State standards & school improvement. *The Elementary School Journal 85*(3):315-336.

Falk, B. (Ed.) (1994). *The Educational Forum*, 59 (1) - Issue on Authentic Assessment.

Falk, B. and Larson, J. (1995). *An invitation to invention: Top-down support for bottom-up reform of assessment in New York State.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Falk, B., MacMurdy, S., & Darling-Hammond, L. (1995). *Taking a different look: How the Primary Language Record supports teaching for diverse learners.* New York: National Center for Restructuring Education, Schools, and Teaching.

Garcia, G.E. & Pearson, P. D. (1994). "Assessment and diversity." In L. Darling-Hammond (Ed.), *Review of Research in Education, 20*, pp. 337-391. Washington, D.C.: American Educational Research Association.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences.* New York: Basic Books.

Glaser, R. & Silver, E. (1994). "Assessment, testing and instruction: Retrospect and prospect." In L. Darling-Hammond (Ed.), *Review of Research in Education, 20*, pp. 393-419. Washington, D.C.: American Educational Research Association.

Kornhaber, M and Gardner, H. (1993). *Varieties of excellence: Identifying and assessing children's talents.* New York: NCREST.

McDonald, J., Smith, S., Turner, D., Finney, M., & Barton, E. (1993). *Graduation by exhibition.* Alexandra, VA: Association for Supervision and Curriculum Development.

Mitchell, R. (1992). *Testing for learning: How new approaches to evaluation can improve American schools.* New York: The Free Press.

Moss, P. A. (1994). Can there be validity without reliability? *Educational Researcher 23*(2), 5-12.

National Association for the Education of Young Children. (January, 1988). NAEYC position statement on developmentally appropriate practice in the primary grades, serving 5 through 8 year olds. *Young Children*, 64-84.

National Forum on Assessment. (1995). *Principles and indicators for student assessment systems.* Cambridge, Mass.: Fairtest.

New York State Curriculum and Assessment Council. (1994) *Learning-centered curriculum and assessment for New York State.* Albany, New York: New York State Education Department.

Perrone, V. (Ed.) (1991). *Expanding student assessment.* Alexandria, VA: ASCD.

Oakes, J. (1985). *Keeping track: How schools structure inequality.* Yale University Press: New Haven.

Piaget, J., & Inhelder, B. (1970). *The science of education and the psychology of the child.* New York: Penguin Books.

Price, J., Schwabacher, S. & Chittenden, E. (1993). *The multiple forms of evidence study.* New York: The National Center for Restructuring Education, Schools, and Teaching (NCREST).

Resnick, L. B. (1987). *Education and learning to think.* Washington, DC: National Academy Press.

Resnick, L. B. (1989). *Tests as standards of achievement in schools.* Paper presented at the Educational Testing Service Conference, The Uses of Standardized Tests in American Education, New York.

Rothman, R. (1995). *Measuring up: Standards, assessment, and school reform.* San Francisco, Calif.: Jossey-Bass.

Shepard, L. (February, 1995). "Using assessment to improve learning." *Educational Leadership*, pp. 38-43.

Sternberg, R.J. (1985). *Beyond IQ.* New York: Cambridge University Press.

Valencia, S., Hiebert, E., and Afflerbach, P. (1994). *Authentic reading assessment: Practices and possibilities.* Newark, Delaware: International Reading Association.

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan, 70,* 703713.

Wiggens, G. (1993). *Assessing student performance: Exploring the purpose and limits of testing.* San Francisco: Jossey-Bass.

Wolf, D.P., (April 1989). Portfolio assessment: Sampling student work. *Educational Leadership,* pp. 35-39.

Wood, D. & Einbender, L. (1995). *An authentic journey: Emergent understandings about authentic assessment and authentic practice.* New York: National Center for Restructuring Education Schools and Teaching.

Principles to Guide Assessment System Design

1. Curriculum, instruction and assessment must be interrelated and interconnected so as to support meaningful student learning.

2. An assessment system must be designed to measure student achievement of standards for learning that delineate what students should know and be able to do as a result of their education.

3. An assessment system should provide multiple forms of evidence of student learning for multiple purposes - but all components of the system should always support student learning.

4. An assessment system should articulate standards without demanding standardization.

5. The assessment system should be built on local involvement.

6. Let the innovators of the system lead.

7. Supports need to be provided to teachers and schools to build their capacities to enact new teaching and assessment practices.

8. How schools are doing should not be judged solely on the basis of student outcomes.

©Falk, Pecheone, and Darling-Hammond, 1996